UNIVERSITY OF BRISTOL

Peer reviewed version

Wipe Scene Change Detection in Video Sequences

W.A.C. Fernando, C.N. Canagarajah, D. R. Bull

Image Communications Group, Centre for Communications Research, University of Bristol,
Merchant Ventures Building, Woodland Road, Bristol BS8 1UB, United Kingdom.
Voice - +44-117-954-5198, Fax - +44-117-954-5206
Email - W.A.C.Fernando@bristol.ac.uk

Abstract

This paper presents a novel algorithm for wipe scene change detection in video sequences. In the proposed scheme, each image in the sequence is mapped to a reduced image. Then we use statistical features and structural properties of the images to identify wipe transition region. Finally, Hough transform is used to analyse the wiping pattern and the direction of wiping. Results show that the algorithm is capable of detecting all wipe regions accurately even when the video sequence contains other special effects.

1. INTRODUCTION

Video is arguably the most popular means of communication and entertainment. Temporal video segmentation, which constitutes the first step in content-based video analysis, refers to breaking the input video into temporal segments with uniform content. Manually partitioning an input video and annotating it with keywords or text is inefficient and inadequate. Therefore, automatic annotation of the input video needs to be developed. Content-based temporal video segmentation is mostly achieved by detection and classification of scene changes (transitions). Basically, transitions can be divided into two categories: abrupt transitions and gradual transitions. Gradual transitions include camera movements such as panning, tilting and zooming, and video editing special effects such as fade-in, fade-out, dissolving and wiping.

Abrupt transitions are very easy to detect, as the two frames are completely uncorrelated. But, gradual transitions are more difficult to detect as the difference between frames corresponding to two successive shots is substantially reduced. Considerable work has been reported on detecting abrupt transitions [1-7]. However, very little effort has been directed toward gradual scene change detection [1-3,8-12].

This paper presents a novel algorithm for wipe scene change detection in video sequences. We exploit statistical features and structural properties of the images and then use Hough transform [13] to identify the wiping pattern and the direction of the wiping. Rest of the paper is organised as follows: Some related work for gradual scene changes are discussed in section 2. Section 3 presents a brief overview of wiping and its application in video production. Section 4 illustrates the proposed algorithm for wipe detection. Results are presented in section 5. Section 6 discusses the conclusions and future work.

2. RELATED WORK

For the detection of gradual scene changes several algorithms have been proposed. Twin comparison method uses the cumulative difference between the frames to detect gradual transitions [1]. This method requires two cut-off thresholds, one higher threshold for detecting abrupt transitions and a lower one for gradual transitions. However, in most of gradual transitions the difference falls below the lower threshold. Therefore, these transitions are not possible to detect with twin comparison. Furthermore, this scheme is not suitable for real time processing or to classify gradual transitions. Zabith et al [2] proposed a feature-based algorithm for detecting and classifying scene breaks. This algorithm requires edge detection in every frame, which is very costly. Another limitation of this scheme is that the edge detection method does not handle rapid changes in overall scene brightness, or scenes, which are very dark or very bright. Furthermore, automatic segmentation and classification is not possible with this scheme.

Alattar proposed statistical feature based approach for wipe detection [10]. This scheme is very sensitive to the type of the video sequence as the algorithm is proposed under a crude approximation for the mean and variance

curves [10]. Furthermore, this cannot identify the nature of wiping such as wiping pattern and the wiping direction. Kim et al presented a wipe detection algorithm based on the visual rhythm [11]. In this scheme, an indexed image is used to find out the visual rhythm. Therefore, each image is represented by a set of lines of the indexed images. Thus performance of this algorithm is dependent on the indexing scheme and the length of the visual rhythm. Furthermore, this cannot be used in real time as it needs a minimum number of frames to evaluate the visual rhythm. Kobla et al discussed the performance of video trails based algorithm [12] to identify video special effects. However, this scheme fails to classify the nature of the special effects, which is essential for video indexing.

3. WIPING

Wipes are widely used in video production to smooth the transitions between two scenes. Wiping is a transition from one scene to another wherein the new scene is revealed by a moving boundary. This moving boundary can be any geometric shape. However in practice this geometric shape is either a line or a set of lines. For an instance horizontal wipe contains a vertical line as its geometric shape of the boundary. An example for horizontal wiping is shown in Figure 1. Table 1 shows line diagrams for some common wiping patterns. According to the geometric shape of the boundary, about 20-30 different moving boundaries are used for wiping in video production. Wipe is considered to be the most difficult gradual scene change to detect due to the sophisticated variation in the moving boundary or the pattern.

Consider a video sequence of length S_E and having a wipe transition from frame W_S to W_E. Then, $W(n)$ can be described as in Equation (1).

$$W(n) = \begin{cases} A(n) & n < W_S \\ P(n) \otimes A(n) + \overline{P}(n) \otimes B(n) & W_S \leq n \leq W_E \\ B(n) & W_E < n \leq S_E \end{cases}$$

$$(1)$$

where, "$\otimes$" denotes element by element matrix multiplication and matrix $P(n)$ generates the wiping pattern.

A(n) - *M*x*N* vector representing the pixel values in frame "*n*" in a video sequence A.
B(n) - *M*x*N* vector representing the pixel values in frame "*n*" in a video sequence B.
P(n) - *M*x*N* vector representing the wiping transition. (Elements of *P(n)* are either "*1*" or "*0*" always).

W(n) - *M*x*N* vector representing the pixel values in frame "*n*" in a video sequence composed from video sequence A and B with wiping.

4. WIPING DETECTION

By subtracting *W(n)* from *W(n-1)*, it is possible to detect wipe transition region. This region moves with the frame number (*n*) according to the wiping pattern. So far it is assumed that both *A(n)* and *B(n)* are fixed frames. However, this is not true in practice. In practice due to motion some movement is possible with both *A(n)* and *B(n)* frames. In these cases, computing pixel-wise luminance difference is not sufficient to detect wipe transition region as pixel-wise difference is highly sensitive to motion within the image. This problem is overcome by dividing each frame into 16x16 pixel blocks and taking its mean and variance to represent each 16x16 block. This block size has been selected in order to detect the minimum number of lines using Hough transform. Therefore, each original frame in the video sequence is mapped into ($M/16, N/16$) reduced image. We defined this reduced index image as the statistical image. Each pixel in the statistical image has two features: mean and variance of the 16x16 blocks of the original image. This is done for all frames in the sequence. Finally, mean square error (MSE) is calculated for corresponding pixels of consecutive statistical images to find out whether a significant change occurred in each block or not. A threshold (T_{MSE}) is used to find out the blocks which have changed during the two consecutive frames. This threshold is an adaptive threshold, which is defined as the mean of the MSEs for all pixels in the statistical image (i.e. T_{MSE} = mean (MSE of all pixels in the statistical image)). Finally, all MSEs are subjected to this threshold T_{MSE} to find out the exact wipe transition region as explained in Equation (2).

$$\begin{aligned} diff\_W(n,i,j) &= 1 & if\ MSE_{n,i,j} > T_{MSE} \\ diff\_W(n,i,j) &= 0 & if\ MSE_{n,i,j} \leq T_{MSE} \end{aligned}$$

$$(2)$$

where, $i = 1: M/16$ and $j = 1: N/16$

Identifying the transition region (in statistical image) is not sufficient to detect wiping automatically. Transition region consists of a single strip or multiple strips and thickness of a strip can be a single line or multiple lines. In practice, wiping transition is achieved over 12-45 frames depending on the image size. Therefore, the thickness of the strips in the statistical image should either be one or two for 176x144 QCIF sequences considered here.

295

The Hough transform is an established technique, which detects a line or a shape by mapping image edge points into a different space called parametric space [13]. Therefore, we can use Hough transform with $diff\_W(n,i,j)$ to identify this transition region whose thickness is a single line or two lines. The number of lines to be detected in parametric space will depend on the block size. If it is small, large number of lines need to be detected in order to identify the wiping patterns. This is due to many blocks changing during two consecutive frames. If the block size is large, it may be difficult to identify the blocks, which have changed during wipe transitions. Therefore, block size is fixed to 16x16 to optimise these two scenarios. Most wiping patterns are generated using one, two or four moving boundaries. Therefore, there are eight lines to be detected at the maximum. This situation arises when four regions are to be detected and the thickness of each region is two lines. Thus, eight highest voted candidates $(V_1 - V_8)$ in the parametric space are analysed.

MSEs are calculated for each pixel in the statistical image and the threshold (T_{MSE}) is used to assign the 2-D binary matrix $diff\_W(n,i,j)$. Then, Hough transformation is applied on $diff\_W(n,i,j)$ to identify the structure of the transition. Highest voted candidates in the parametric space $(V_1 - V_8)$ are analysed to identify the four lines and calculate the average gradient. If it is not possible to identify four lines, then algorithm tries to identify two lines and the average gradient is assigned accordingly. Otherwise, it identifies a single line and assigns the average gradient as gradient of V_1 or V_1 and V_2 depending on the thickness of the strips. Average gradient should be a constant for a wipe transition. Finally, value of the average gradient and the number of lines reveal the wiping pattern. Having identified a wipe transition, the next step is to identify the wiping direction. Wipe direction is dependent on the constants of lines. Therefore, direction of the wiping pattern is identified by checking the variation of the constants of $V_1 - V_8$. If the thickness of the stripes is two lines, then the maximum constant (out of two lines) is considered to represent a single strip. Following steps summarise the complete algorithm.

Step 1: Compute the MSEs for each pixel in the statistical image.

Step 2: Threshold the calculated MSE values with T_{MSE} and assigned $diff\_W(n,i,j)$.

Step 3: Apply Hough transformation on $diff\_W(n,i,j)$

Step 4: Check the status of $V_1 - V_8$ to identify four lines involved and calculate the average gradient. If it is not possible to identify four lines, then identify two lines and the average gradient is assigned as previously. Otherwise, identify a single line and assign the average gradient as gradient of V_1 or V_1 and V_2. Average gradient should be a constant for a wipe transition. Finally, value of the average gradient and the number of lines reveal the wiping pattern.

Step 5: If step 4 is satisfied then check the variation of the constant of $V_1 - V_8$ to find out the direction of wiping.

Step 6: Back to step 1.

5. RESULTS

Consider a test sequence, which contains vertical wiping, to describe the performance of the above algorithm. Figure 2 and Figure 3 show the average the gradient and the constant respectively. Wiping pattern is identified from the gradient curve and the direction of wiping pattern is identified from the variation of the constant curve. From Figure 2 it is clear that the wiping pattern is vertical since the average gradient is 180^0. Since constant is increasing (during the period of wiping) with the frame number, wiping direction should be forward. Therefore, forward vertical wiping pattern is identified from 42nd frame to 76th frame.

Table 2 shows the summarised results of the proposed algorithm with the sequence 1 and sequence 2. These results show that the algorithm is capable of detecting all wipe regions accurately even when the video sequence contains other special effects or camera effects. There are two main advantages of this algorithm: no external thresholds are involved and detailed classification of the wiping patterns is possible. Therefore, the proposed algorithm can be used to detect wipe regions in video sequences.

6. CONCLUSIONS

In this paper, we have presented a novel algorithm for wipe scene change detection in video sequences. We exploited the statistical features and structural properties of the images and then used Hough transform to identify the wiping pattern and the direction of the wiping. Results show that the algorithm is capable of detecting all wipe regions accurately even when the video sequence contains other special effects like fading, dissolving, panning etc,.

Therefore, the proposed algorithm can be used in uncompressed video to detect wipe regions with a very high reliability. Further work is required to extend this algorithm for compressed video.

ACKNOWLEDGEMENTS

References

1. Nagasaka, and Y. Tanaka, *"Automatic Video Indexing and Full-Video Search for Object Appearances,"* Visual Database Systems II, Eds.E.Kunth, and L.M. Wegner, Elsevier Science Publishers B.V., IFIP, pp. 113-127,1992.

2. Zabith, R., Miller, J., and Mai, K., *"Feature-Based Algorithms for Detecting and Classifying Scene Breaks,"* 4th ACM International Conference on Multimedia, San Francisco, California, November-1995.

3. Yeo, B.L., *"Rapid Scene Analysis on Compressed Video"*, IEEE Transactions on Circuits and Systems for video technology, Vol. 5, No 6, pp. 533-544, December-1995.

4. Zhang, H.J., *" Automatic Partitioning of Full-Motion Video"*, ACM/Springer Multimedia Systems, Vol.1, No.1, pp. 10-28, 1993.

5. Shin, T. et. al, *"Hierarchical scene change detection in an MPEG-2 compressed video sequence"*, Proceedings - IEEE International Symposium on Circuits and Systems, Vol.4, pp. 253-256, 1998.

6. Fernando, W.A.C., Canagarajah, C.N., Bull, D. R., *"Sudden Scene Change Detection in MPEG-2 Video Sequences,"* Paper Number - 13, Proceedings - International Workshop on Multimedia Signal Processing, 1999.

7. Fernando, W.A.C., Canagarajah, C.N., Bull, D. R., *"DFD Based Scene Segmentation For H.263 Video Sequences"*, Paper Number-747, Proceedings - IEEE International Symposium on Circuits and Systems, 1999.

8. Fernando, W.A.C., Canagarajah, C.N., Bull, D. R., *"Automatic Detection of Fade-in and Fade-out in Video Sequences,"* Paper Number-748, Proceedings - IEEE International Symposium on Circuits and Systems, 1999.

9. Fernando, W.A.C., Canagarajah, C.N., Bull, D. R., *"Video Segmentation and Classification for Content Based Storage and Retrieval Using Motion Vectors,"* pp. 687-698, Storage and Retrieval for Image and Video Databases VII - SPIE, San Jose, California, USA, 1999.

10. Alattar, A. M., *"Wipe Scene Change Detector For Segmenting Uncompressed Video Sequences,"* Proceedings - IEEE International Symposium on Circuits and Systems, Vol.4, pp. 249-252, 1998.

11. Kim, H., Park, S.J, Kim, W.M., Song, M.H., *"Processing of Partial Video Data for Detection of Wipes"*, pp.280-289, Storage and Retrieval for Image and Video Databases VII - SPIE, San Jose, California, USA, 1999.

12. Kobla, V., DeMenthon, D., Doermann, D., *" Special Effect Edit Detection Using Video Trails: a Comparison with Existing Techniques,"* pp.302-313, Storage and Retrieval for Image and Video Databases VII - SPIE, San Jose, California, USA, 1999.

13. Dana H. Ballard, Christopher B. Brown, *"Computer Vision,"* Prentice-Hall, 1982.
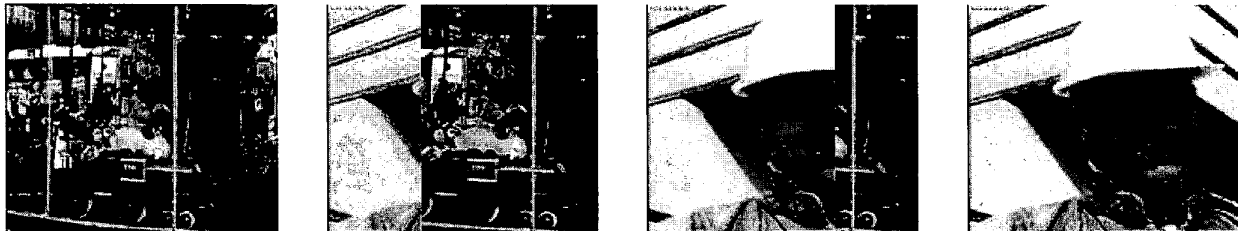

Figure 1: Horizontal wiping

Notation	Wiping Pattern	Average gradient
W-1		90^0
W-2		90^0
W-3		180^0
W-4		180^0
W-5		45^0
W-6		45^0
W-7		135^0
W-8		135^0
W-9		135^0
W-10		135^0
W-11		90^0
W-12		90^0
W-13		180^0
W-14		180^0

Table 1: Common wiping patterns

Sequence 1   Length of 800 frames and contains eight wipe regions. This sequence does not contain any other gradual scene changes.

Sequence 2   Length of 1500 frames and contains twelve wipe regions, five sudden scene changes, and several other special effects like fade-in, fade-out, dissolving and camera movements such as zoom-in, zoom-out, panning and tilting.

Sequence	Actual wipe region	Detected wipe region	Nature of wiping
Sequence 1	25-54	25-54	W-1
	98-144	98-144	W-4
	203-237	203-237	W-2
	340-378	340-378	W-7
	397-442	397-442	W-11
	539-574	539-574	W-12
	623-637	623-637	W-14
	688-725	688-725	W-1
Sequence 2	23-53	23-53	W-1
	102-131	102-131	W-3
	176-212	176-211	W-9
	258-295	258-295	W-2
	387-414	387-414	W-14
	465-507	465-507	W-10
	569-607	569-607	W-4
	780-827	780-827	W-1
	920-964	920-964	W-3
	1090-1124	1090-1124	W-8
	1223-1256	1223-1256	W-8
	1402-1448	1402-1447	W-10

Table 2: Summarised results for wiping in video sequences with the proposed algorithm
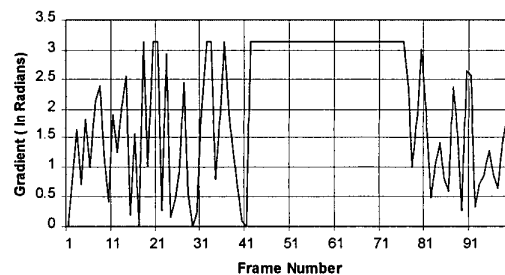

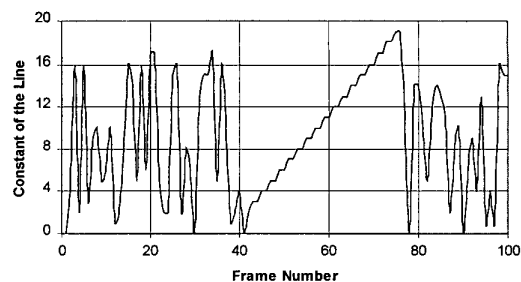Figure 2: Average gradient of highest voted candidate(s) in parametric space

Figure 3: Constant of the highest voted candidate in parametric space

298